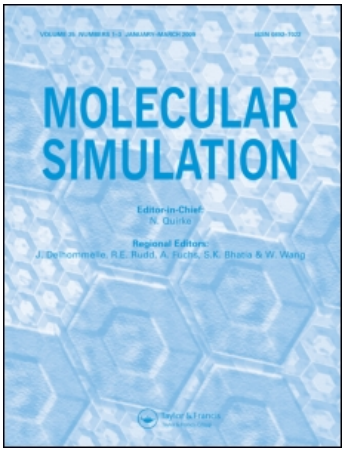


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

A CHARMm Based Force Field for Carbohydrates Using the CHEAT Approach: Carbohydrate Hydroxyl Groups Represented by Extended Atoms

Peter D. J. Grootenhuis[a]; Cornelis A. G. Haasnoot[a]
[a] Department of Computational Medicinal Chemistry, AKZO Pharma Division, Organon International bv, Oss, The Netherlands

To cite this Article Grootenhuis, Peter D. J. and Haasnoot, Cornelis A. G.(1993) 'A CHARMm Based Force Field for Carbohydrates Using the CHEAT Approach: Carbohydrate Hydroxyl Groups Represented by Extended Atoms', Molecular Simulation, 10: 2, 75 — 95
To link to this Article: DOI: 10.1080/08927029308022160
URL: http://dx.doi.org/10.1080/08927029308022160

Molecular Simulation, 1993, Vol. 10(2-6), pp. 75-95
Reprints available directly from the publisher
Photocopying permitted by license only

A CHARMm BASED FORCE FIELD FOR CARBOHYDRATES USING THE CHEAT APPROACH: CARBOHYDRATE HYDROXYL GROUPS REPRESENTED BY EXTENDED ATOMS

PETER D.J. GROOTENHUIS* and CORNELIS A.G. HAASNOOT

Department of Computational Medicinal Chemistry, Organon International bv, AKZO Pharma Division, P.O. Box 20, 5340 BH Oss, The Netherlands

(Received December 22, 1992, accepted December 23, 1992)

Computational studies of carbohydrates that do not consider explicit solvent molecules suffer from the strong tendency of the carbohydrate pendant hydroxyl groups to form intramolecular hydrogen bonds that are unlikely to be present in protic media. In this paper a novel approach towards molecular modelling of carbohydrates is described. The average effect of intra- and intermolecular hydrogen bonding is introduced into the potential energy function by adding a new (extended) atom type representing a carbohydrate hydroxyl group to the CHARMm force field; we coin the name CHEAT (*C*arbohydrate *H*ydroxyls represented by *E*xtended *A*Toms) for the resulting force field. As a training set for the parametrisation of CHEAT we used ethylene glycol, 10 cyclohexanols, 5 inositols, and 12 glycopyranoses for which in total 64 conformational energy differences were estimated using a set of steric interaction energies between hydroxyl and/or methyl groups on six-membered ring compounds as derived by Angyal (*Angew. Chem.*, **81**, 172-182, (1969)). The root-mean-square deviation between the estimated energy differences and the corresponding values obtained by our CHEAT approach amounts to 0.37 kcal/mol (n = 64). The CHEAT approach, which is claimed to calculate aqueous state compatible energetical and conformational properties of carbohydrates, is computationally very efficient and facilitates the calculation of nanosecond range MD trajectories as well as systematic conformational searches of oligosaccharides.

KEY WORDS: Carbohydrates; force field; molecular mechanics, conformational analysis; solvation; hydrogen bonding

INTRODUCTION

Parallel to the renewed interest in carbohydrates as an important class of biomolecules [1], the development of molecular modelling methods for carbohydrates has been receiving a growing amount of attention during the last decade [2,3,4,5]. Over the years, many computational approaches towards studying the conformational and energetical aspects of carbohydrates have been deployed. Inspection of the literature shows that one of the most widely used approaches is the so-called HSEA (*H*ard *S*phere *E*xo-*A*nomeric) method [6], however, evidence exists that this method leads to a too highly restricted conformational flexibility [7]. Examples of extensions of the HSEA approach are the HEAH (hydrogen bond potential added), NBEA, PF, and PFOS methods [7]. Using the Consistent Force Field methodology, Rasmussen developed the PEF422 force field [8]; Brady [9] implemented the latter force field in CHARMM [10] for use in several carbohydrate studies. Others have tried to modify established general valence force fields, in most cases by adding

parameters that mimic the anomeric effect. In this way MM2CARB [11] (based on MM2 [12]) was found to give good agreement with experimental results [7]. Homans [13] and Scarsdale et al. [14] modified the AMBER [15] force field for carbohydrates. In the GEGOP (*Ge*ometry of *Glyco**p**e*ptides) program an extension of the HSEA procedure, designated GESA (*Ge*ometry of *Sa*ccharides), is used for carbohydrates and the ECEPP/2 force field for peptides [16]. Also unmodified general purpose force fields like GROMOS and MM3 have been used for carbohydrate modelling studies [17,18]. Very recently, Imberty et al. [19] added force field parameters for simulating carbohydrates to the TRIPOS force field [20] implemented in the SYBYL molecular graphics package.

Apart from the "classical" problems [21] in force field calculations (such as the inability of energy minimisation methods to locate global minima, the rather crude approximations used in most methods, i.e. no polarisation or many-body interactions, etc.), all of the above mentioned methods suffer from a number of problems that are directly related to the nature of carbohydrates. The most serious problem regards the phenomenon of hydrogen bond formation in *in vacuo* calculations, i.e. in the absence of explicit solvent molecules. In that case, the pendant hydroxyl groups of carbohydrates start forming intramolecular hydrogen bonds with the adjacent hydroxyl groups upon energy minimisation and it turns out that the energies associated with these hydrogen bonds dominate the total energy of the system. Exemplary is the case of β-D-glucose: by systematically generating the three staggered orientations of each pendant hydroxyl group of glucose, i.e. 3^5 ($= 243$) conformations, it was found [22] that the final steric energies of these conformations as predicted by CHARMm [23] varied over a range of 20 kcal/mol! Of course, the lowest energy conformer is a conformation in which the hydroxyl groups are engaged in an intricate hydrogen bonding scheme, but it is hard to predict in advance which scheme yields the lowest energy conformation. For oligosaccharides the problem is carried to extremes as the number of possible conformations grows exponentially [24]. In practice, a systematic search is impossible for tetrasaccharides [25] and up.

But apart from this logistic argument, there is also the more basic question on how relevant the thus found global minimum energy conformation is when it is realised that one is seldom interested in the gas-phase conformation of (oligo-) saccharides. Often it is tacitly assumed that the calculated (gas-phase) conformation also represents the solvated conformation, but in the case of carbohydrates this seems a perilous assumption as it is long suspected that intramolecular hydrogen bonding in carbohydrates is relegated to a minor role in hydroxylic solvents [26, 27,28,29,30]. This is also illustrated by molecular dynamics simulations in which solvent molecules are explicitly included in the calculation. From such simulations the balance between intra- and intermolecular hydrogen bonding can in principle be assessed, assuming that the carbohydrate and solvent parameters are properly balanced. Starting with glucose in the aforementioned lowest energy (gas-phase) conformation but now surrounded by a droplet of water molecules, a molecular dynamics simulation [22] demonstrates that already after a few ps the initial intramolecular hydrogen bonding array of the carbohydrate is gravely perturbed and instead water molecules act as hydrogen bonding partners for the pendant hydroxyl groups.

Taken together, it is not surprising that French and Brady [31] concluded that ". . . uncertainty regarding hydrogen bonds overshadows many other types of errors in modelling studies (of carbohydrates)". Molecular dynamics simulations which

include solvent molecules can alleviate [32] this problem, however, such calculations take vast amounts of cpu time on high-speed computers and sampling errors, etc., may still lead to inconsistent results [33]. Therefore, such an approach seems to be impractical for relatively simple conformational studies needed to interpret NMR data of oligosaccharides, etc.

In search of an alternative method we explored the possibility to circumvent the above mentioned problems by simply removing the cause, viz. the hydroxyl proton. *In computro* this can be realised by replacing the hydroxyl groups of the carbohydrate molecules by a specially parametrized extended atom type. It is noted in passing that the concept of "extended" (also called "united") atoms, in which two or more atoms are represented by a single atom type, has been applied very effectively in many force fields, albeit in nearly all cases for apolar functionalities, e.g. methylene, or methyl groups. Therefore, it is likely that such an extended "hydroxyl" atom type can be added to any general valence force field, but in the present paper we restrict ourselves to CHARMm. Since we aim for a good correspondence between "gas-phase" calculations (i.e. no explicit solvent molecules included) and solution conformations it is essential that the parameters of this additional atom type mimic the time-averaged effect of both intra- and intermolecular hydrogen bonding on the energetics of the carbohydrate molecules at hand [34].

There are a number of potential advantages to this approach which we dubbed CHEAT (*C*arbohydrate *H*ydroxyls represented by *E*xtended *AT*oms). Firstly, the introduction of a new atom type allows the inclusion of the anomeric [35,36] and gauche [37,38] effects without interfering with the rest of the basic (CHARMm) force field. Secondly, the number of atoms and orientations is reduced, resulting in a simplified description of the carbohydrates and a concomitant speed-up of the calculation. Thirdly, no explicit solvent molecules have to be included in the calculation, thereby allowing fast scanning of conformational space accessible to an oligosaccharide.

In the present paper it is reported how the proposed extended "hydroxyl" atom type was parametrised. Next, the resulting CHEAT force field is evaluated in terms of energetics and geometries. Finally, the effectiveness of the CHEAT approach will be exemplified by the generation of a (Φ, Ψ) iso-energy contour map for cellobiose and by several 1 ns molecular dynamics studies of D-glucose.

COMPUTATIONAL METHODS

Molecular Mechanics

All force field calculations were carried out using CHARMm version 21.2 present in the Quanta/CHARMm molecular modelling package version 3.0 [39]. A distance dependent dielectric constant ($\epsilon = r$) and no cut-off for the nonbonded interactions was used. Conjugate gradient energy minimisations were continued until the root-mean-square (rms) energy gradient was less than 0.001 kcal/mol.Å. Unless otherwise described the atomic charges and parameters present in Quanta's POLYSACH30.RTF and PARM30.PRM files served as a basis for the CHEAT force field development described below.

Molecular Dynamics

The molecular dynamics protocol consisted of three parts: for the heating to 300 K, subsequent equilibration and dynamics, trajectories of 1 ps, 10 ps, and 1000 ps, respectively, were calculated using the Verlet algorithm. During the last trajectory, which was used for further analyses, the coordinates were saved every 0.1 ps. The SHAKE algorithm [40] was used for the C-H bonds, thus allowing a time step of 0.001 ps.

Conformational Maps

For the conformational map of cellobiose the two dihedral angles that define the glycosidic bond, i.e. H1-C1-04′-C4′ (Φ) and C1-04′-C4′-H4′ (Ψ), were varied from $-180°$ to 180° using 20° increments. At each grid point an energy minimisation was carried out with the dihedral angles Φ and Ψ constrained by applying a force constant of 1000 kcal/mol.rad^2. The energy minimisation protocol consisted of 50 steps of steepest descent, followed by a conjugate gradient minimisation until the energy gradient was less than 0.001 kcal/mol.Å. After convergence the constraints were removed and the total energy was recalculated. Finally, an iso-energy contour map was drawn as a function of Φ and Ψ.

PARAMETRISATION OF CHEAT

General Considerations

The basic idea behind the CHEAT approach is the introduction of the average effect of intra- and intermolecular hydrogen bonding into the potential energy function by adding extended atoms for the carbohydrate hydroxyl groups to an existing general valence force field (in the present study: CHARMm). For this new extended atom type (designated OT1E) all force field parameters are then to be supplied. Ideally, no changes are introduced to the CHARMm force field, only new values are added. However, it is noted here already that this premise could not be maintained and therefore two minor but important adaptations of the basic CHARMm force field had to be introduced (*vide infra*).

Whenever plausible, parameter values were copied from corresponding CHARMm force field atom types. However, for some missing parameters new values had to be derived. For this we used a training set of molecules for which conformational energy differences were calculated as a function of the parameters to be derived. By performing a grid search in which these parameters were systematically varied, parameter values were obtained that gave an optimal fit between calculated and experimental energy differences.

Training Set

In order to gauge missing parameters of the extended "hydroxyl" atom type to be added to the CHARMm force field, a consistent set of reference data is prerequisite. Since systematic thermodynamic studies on e.g. simple sugars in solution

Table 1 Conformational energy differences (kcal/mol) between axially and equatorially substituted cyclohexane derivatives (in the chair conformation) and *gauche* vs. *anti ethanediol.*

Compound	*Experimental*[a]	*CHEAT*[b]
cyclohexanol (*a* ⇄ *e*)	0.90	0.83
1,2-cyclohexadiol (*a,a* ⇄ *e,e*)	1.45	1.72
1,3-cyclohexadiol (*a,a* ⇄ *e,e*)	2.40	2.23
1,4-cyclohexadiol (*a,a* ⇄ *e,e*)	1.80	1.63
1,2-trans-methylcyclohexanol (*a,a* ⇄ *e,e*)	2.25	1.93
1,2-cis-methylcyclohexanol (*a,e* ⇄ *e,a*)	0.90	0.79
1,3-methylcyclohexanol (*a,a* ⇄ *e,e*)	3.85	3.60
1,3,5-methylcyclohexadiol (*a,a,a* ⇄ *e,e,e*)	6.50	5.81
1,2,3-cyclohexatriol (*a,a,a* ⇄ *e,e,e*)	2.60	2.92
1,3,5-cyclohexatriol (*a,a,a* ⇄ *e,e,e*)	4.50	4.17
(+) inositol (*e,e,a,a,a,a* ⇄ *a,a,e,e,e,e*)	2.30	2.32
(epi) inositol (*e,a,a,a,e,a*) ⇄ *a,e,e,e,a,e*	2.30	2.45
(myo) inositol (*e,a,a,a,a,a* ⇄ *a,e,e,e,e,e*)	4.60	4.77
(neo) inositol (*e,a,a,e,a,a* ⇄ *a,e,e,a,e,e*)	2.30	2.68
(scyllo) inositol (a,a,a,a,a,a ⇄ *e,e,e,e,e,e*)	6.90	6.79
ethanediol (*anti* ⇄ *gauche*)	0.70[c]	0.19

[a] Estimated using Angyal's scheme [34].
[b] Present paper.
[c] Refs. 35 and 36.

are lacking we had course to the generalised steric interaction energy scheme proposed by Angyal [41]. The latter scheme allows one to estimate the difference in conformational energy between six-membered ring compounds with substituents in axial and in equatorial positions, respectively. There are several advantages in, using Angyal's scheme: 1) energy differences can be calculated for an arbitrary substituted cyclohexane or hexapyranose ring system; 2) the energy differences thus obtained reflect the situation in aqueous solution at 298 K, i.e. the state we for to reproduce by our *in vacuo* calculations; 3) the estimated conformational energy data are not obscured by entropy of mixing, etc.

Several classes of molecules were used in order to optimise missing parameters for the new OT1E extended atom type. The first set of molecules consisted of 12 hydroxy and/or methyl substituted cyclohexanes together with five inositols and 1,2-ethanediol (Table 1); the second set consisted of the 12 aldopyranoses (Table 2). With the exception of ethanediol (where the experimental [42,43] energy difference between the anti and gauche conformer was used), the differences in steric interaction energies as calculated by means of Angyal's scheme for the discernible chair conformers were used to calibrate the parameters of the extended hydroxyl atom type. As indicated in Table 2, for every aldopyranose four interaction energy differences between the α- and β-anomers in both the $^4C_1$ and the $^1C_4$ conformation [44] were calculated. In this way a total of 64 energy differences estimated for a training set of 28 molecules were available for the parametrization of the CHEAT force field.

Table 2 Conformational energy differences (kcal/mol) between the α- and β-anomers of the $^4C_1$ and $^1C_4$ conformations of the D-aldopyranoses as estimated from Angyal's steric interaction energy scheme [41] and calculated using CHEAT.

	$^4C_1(\alpha)$-$^4C_1(\beta)$		$^1C_4(\alpha)$-$^1C_4(\beta)$		$^1C_4(\alpha)$-$^4C_1(\alpha)$		$^1C_4(\beta)$-$^4C_1(\beta)$	
Compound	*Angyal*	*CHEAT*	*Angyal*	*CHEAT*	*Angyal*	*CHEAT*	*Angyal*	*CHEAT*
allose	0.95	0.89	−0.70	−1.17	1.45	1.52	3.10	3.58
altrose	0.30	0.42	−1.50	−1.47	0.20	0.28	2.00	2.17
galactose	0.35	0.06	−1.45	−1.82	3.45	2.54	5.25	4.63
glucose	0.35	0.21	−1.45	−1.76	4.15	3.63	5.95	5.60
gulose	0.95	0.73	−0.70	−1.17	0.75	0.60	2.40	2.50
idose	0.30	0.21	−1.50	−1.62	−0.50	−1.05	1.30	0.78
mannose	−0.45	−0.12	−2.10	−2.19	3.05	2.56	4.70	4.63
talose	−0.45	−0.28	−2.10	−2.27	2.35	1.15	4.00	3.14
arabinose	0.30	0.27	−0.35	0.03	−1.15	−1.04	−0.50	−0.80
lyxose	−0.45	−0.25	−0.95	−0.64	0.55	0.74	1.05	1.12
ribose	0.95	0.74	0.45	0.36	0.10	−0.50	0.60	−0.38
xylose	0.35	0.09	−0.30	−0.14	1.65	1.74	2.30	1.96

Extended Atom Type OT1E Parameters

The bond and angle parameters for the new extended atom type OT1E were directly copied from the corresponding CHARMm parameters given for a standard hydroxyl oxygen (atom type OT). The same holds for the dihedral parameters. However, for two dihedral angles that involve the OT1E atom type, i.e. the OT1E-CT-CT-OT1E and OT1E-CT-OE-CT dihedrals, new force constants had to be derived (*vide infra*). It turned out to be necessary to introduce 2-fold and 3-fold periodicities simultaneously for each dihedral in order to account for the gauche and anomeric effects.

Also for the 6-12 Lennard-Jones parameters Emin and R* of the OT1E atom type new values had to be determined. In CHARMm separate Lennard-Jones parameters can be used for the 1-4 interactions. The Emin value for the 1-4 interactions of OT1E was fixed at 0.10, being equal to the value for the extended atom type CH1E (the extended atom representation for a carbon atom with one hydrogen). The R* for the 1-4 interactions value was optimised as will be described below. As a consequence of using the new extended atom, also new partial atomic charges had to be derived for the OT1E atoms and the adjacent carbon atoms. The partial charges for the hydrogens and C5 atom were copied from the standard glucose residue available in CHARMm. It follows that in total values for nine new force field parameters had to be derived.

Parametrisation and Results

The values of the nine missing parameters of the hydroxyl extended atom type were optimised by carrying out several grid searches. In the grid searches the values for the missing parameters were varied between what we considered to be reasonable extremes, first using large steps and followed by a finer grid. For each combination of parameter values the relevant conformers of the 28 molecules given in Tables 1 and 2 were energy minimised so that the 64 energy differences in the training set of 28 molecules could be calculated and compared with the corresponding training

set values. Because this procedure is extremely computation-intensive the grid search was split up in several smaller 3- and 4-dimensional grid searches. Throughout the grid searches the data were analysed with the SAS package, vs. 6.03 (SAS Institute Inc., Cary, NC). The sum of squares and rms deviation between the experimental and calculated energy differences served as measures for the goodness-of-fit. Upon executing the grid searches, it became clear that some parameters are strongly correlated so that a number of virtually equivalent solutions exist. In these cases we decided to select values for the variables that are best in line with the other parameters in the CHARMm parameter set.

In a first attempt, we tried to generate the new extended atom parameters without modifying the original CHARMm force field. However, under the latter constraint it turned out to be impossible to arrive at an acceptable correspondence between calculated and Angyal derived energy differences when all 28 molecules of the training set were taken into account. In fact, the best attainable result was obtained when only the energy data for aldopyranoses and inositols were taken into consideration: in that case the rms difference between calculated and Angyal derived data could be optimised to 0.68 kcal/mol (N = 53). Scrutiny of the results of the latter optimisation, however, revealed a remarkable difference between the data for the C5-sugars (which lack an exocyclic hydroxymethylene moiety at the 5-position) and the C6-sugars. It was found that the rms difference between calculated and Angyal derived data for $^4C_1 - {}^1C_4$ interconversions was significantly better for C5-sugars (rms difference 0.51 kcal/mol, N = 13) than for C6-sugars (rms difference 1.12 kcal/mol, N = 16). This finding could be reduced to a basic flaw in the original CHARMm parameter set: calculation of the equatorial and axial conformer of methylcyclohexane showed that the so-called A-value predicted by CHARMm amounts to 0.9 kcal/mole, i.e. considerably lower than the experimental value of 1.8 kcal/mol [45].

Therefore, despite our initial intentions not to introduce any changes to the CHARMm parameter set, it turned out to be necessary to adapt two parameters in order to correct for some serious shortcomings in the basic CHARMm force field. The 10 first change comprises the torsion barrier of the X-CT-CT-X dihedral angle, which was lowered from 1.6 kcal/mol to 1.0 kcal/mol. Moreover, the R* value of the CT atom type was increased from 1.8 Å to 2.15 Å. These changes to the CHARMm parameter set result in an improved conformational energy prediction for two basic molecules, viz. ethane and methylcyclohexane. Now, the ethane barrier amounts to 2.9 kcal/mol (versus 4.1 kcal/mol using the original CHARMm field) and the difference between axially and equatorially substituted methylcyclohexane is 1.8 kcal/mol; both new values are in excellent agreement with experiment [46].

A second grid search using the modified CHARMm force field resulted in a set of OT1E parameters for which the rms deviation between the Angyal and CHEAT calculated energy differences amounts to 0.37 kcal/mol for all 64 energy differences pertaining to the training set. For the 24 α/β-anomeric equilibria considered, the rms deviation is even smaller, viz. 0.25 kcal/mol. Notwithstandingly, the imbalance between the C5- and C6-sugars persists to some extent: also for this parametrisation the C5-sugars are calculated more accurately (rms difference 0.21 kcal/mol, N = 13) than the C6-sugars (rms difference 0.59 kcal/mol, N = 16). Although the latter finding demonstrates that our ad hoc adaptations of the basic CHARMm force field were only partially successful, the obtained parameter, set is the best

we can do next to fully reparametrising the basic CHARMm force field. We will refer to the thus obtained set as the CHEAT force field. A full account of the force field parameters needed for the calculation of carbohydrates using the CHEAT approach is given in the Appendix.

Pyranose Ring Conformations

As was described above, the parametrisation of the CHEAT force field aimed for an optimal match between calculated and experimental values of conformational energy differences. It follows that structural criteria were relegated to a minor role in this parametrisation process: apart from monitoring the calculated structures for complying with the conformational equilibria under study, no attention was paid to detailed structural elements such as bond lengths, bond angles, etc. Therefore, an important test for the CHEAT force field is a check on its ability to reproduce key structural features in monosaccharides.

We focused on the six-membered ring conformations in four monosaccharides, viz. α- and β-D-glucopyranose and α-/β-D-mannopyranose. The Cambridge Structural Datafile (CSD) [47,48] was scanned for these (sub-)structures and average values of bond lengths, bond angles and torsion angles occurring in these four saccharide units were determined. The latter values served as experimental reference points in our evaluation. As a second reference point the corresponding structural data were determined for the four above mentioned structures calculated using the standard CHARMm 21.2 force field. Scrutiny of the data thus obtained showed that both the CHEAT and the standard CHARMm force field reproduced the averaged experimental data virtually equally well. For example, the rms difference between the twelve bond distances between the heavy atoms in α-D-glucopyranose calculated by means of the CHEAT force field on the one hand and the corresponding averaged experimental bond distances on the other hand amounts to 0.014 Å; the corresponding rms difference between the bond lengths calculated by CHARMm and the averaged experimental bond lengths aggregates to 0.011 Å. Both rms differences are comparable to the standard deviations calculated for the experimental averages (0.008–0.014 Å). Comparison of 17 bond angles between heavy atoms in α-D-glucopyranose calculated by CHEAT and the corresponding X-ray averages arrives at an rms difference of 1.7°, whereas for the structure predicted by CHARMm the corresponding rms difference is 2.2°; both rms differences fall within the range of standard deviations calculated for the X-ray averages: 0.9°–3.5°. Finally, the torsion angles involving heavy atoms show an rms difference between the torsions taken from the CHEAT structure versus the X-ray torsion averages of 4.3°, for the CHARMm structure a corresponding rms difference of 8.1° is noted (standard deviations in the X-ray torsion averages vary from 4.0°–8.5°).

From a statistical point of view the correspondence between the calculated and the (averaged) X-ray structural features is satisfactory. However, especially the above comparison of "isolated" torsion angles cannot reveal interrelated deviations which may affect overall conformational features. Therefore, the calculated and experimental conformations of the six-membered rings in α/β-D-glucose and α/β-D-mannose were characterised in terms of Cremer-Pople (CP) puckering coordinates [49]. For a full comparison the reader is referred to Table 3, here suffice it to note that compared to the averaged X-ray CP-parameters CHARMm in general

Table 3 Cremer-Pople [49] puckering coordinates (q_2, Φ_2, q_3) calculated for the α- and β-anomers of glucose and mannose from sets of X-ray structures (taken from the CSD [47,48]) and structures calculated by means of CHARMm 21.2 (all atom) and CHEAT.

Compound	*Method*	q_2(Å)	Φ_2	q_3(Å)
α-D-Glucose	⟨X-ray⟩ (n = 42)	0.063 (0.055)	133.5° (92.7°)	−0.556 (0.033)
	CHARMm 21.2	0.048	328.6	−0.572
	CHEAT	0.031	72.2	−0.520
β-D-Glucose	⟨X-ray⟩ (n = 46)	0.070 (0.027)	130.4° (51.5°)	−0.578 (0.018)
	CHARMm 21.2	0.052	15.8°	−0.563
	CHEAT	0.060	136.8°	−0.535
α-D-Mannose	⟨X-ray⟩ (n = 23)	0.096 (0.084)	129.9° (120.8°)	−0.559 (0.043)
	CHARMm 21.2	0.025	330.6°	−0.582
	CHEAT	0.035	74.6°	−0.518
β-D-Mannose	⟨X-ray⟩ (n = 6)	0.062 (0.028)	150.6° (64.7°)	−0.563 (0.035)
	CHARMm 21.2	0.052	15.8°	−0.563
	CHEAT	0.060	136.8°	−0.536

overestimates the overall puckering amplitude of the six-membered rings studied, whereas CHEAT somewhat underestimates them. On basis of the Φ_2-values summarised in Table 3 it is concluded that the local flattening in the calculated chair conformations is better predicted by the CHEAT force field than by the original CHARMm 21.2 field.

APPLICATIONS OF THE CHEAT FORCE FIELD

Conformational Analysis of Cellobiose

Carrying out a systematic conformational analysis of a disaccharide by means of computational chemistry tends to be rather cumbersome when an all atom description of the carbohydrate is used [50]. The generation of iso-energy Φ, Ψ-contour maps requires a large number of disaccharide conformations to be calculated *in vacuo* and, as was argued in the Introduction, the positions of the exocyclic hydroxyl groups affect the calculated energy values to a large extent. Forced by the potential size of the computational problem (*vide supra*), most authors limit themselves to exploring only the two intramolecular co-operative hydrogen bonding schemes, oriented "clockwise" and "anticlockwise" about the ring, as starting points for their conformational analyses. This selection is rather arbritary and, moreover, during the systematic grid search about the two dihedrals Φ and Ψ that define the glycosidic bond, intramolecular hydrogen bonds between the two saccharide moieties may be formed leading to local energy minima that are not substantiated by experiment.

The CHEAT approach to carbohydrate modelling is geared to dealing with such conformational analyses of oligosaccharides. As an illustration, the iso-energy Φ, Ψ-contour map of a (1 → 4)-D-glucopyranosyl disaccharide, i.e. β-cellobiose (cf. Figure 1) was calculated using CHEAT; such a conformational map calculation using a 20° grid only takes approximately one cpu hour on a Silicon Graphics 4D/35TG computer. As can be gleaned from Figure 1a, the contour map is typical for a flexible-residue approach with relatively low energy barriers separating the (local) minimum energy conformations accessible to the disaccharide. In fact, the

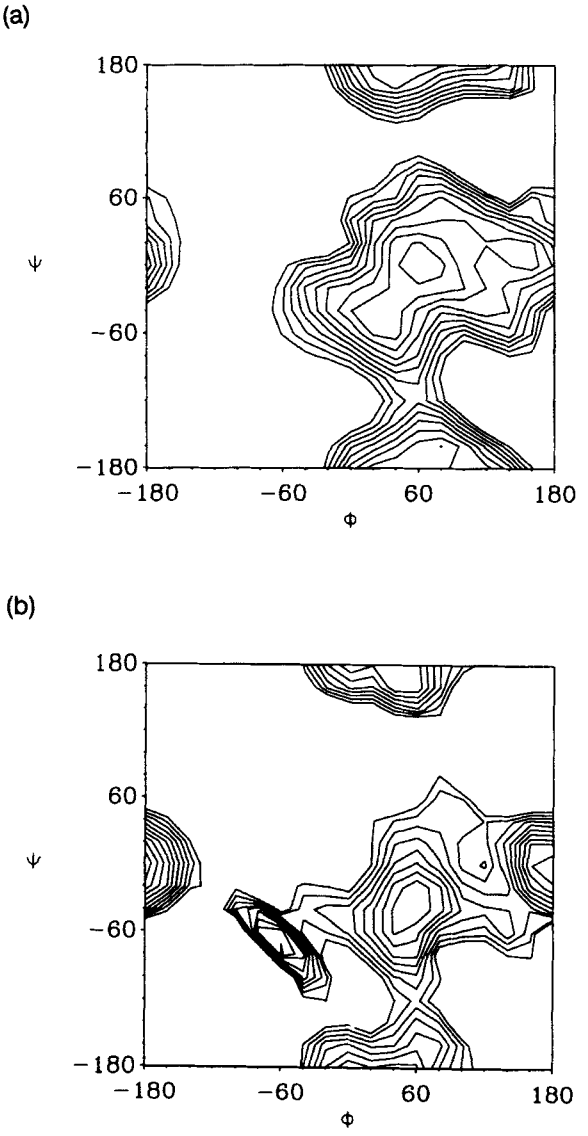
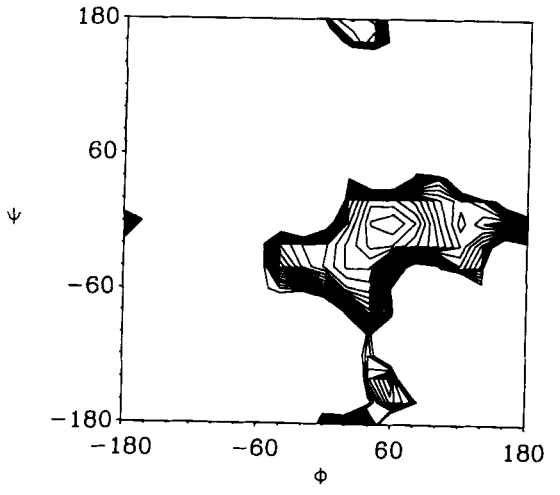


Figure 1 Iso-energy (Φ, Ψ) maps for cellobiose calculated using the (a), all atom CHARMm (b) and HSEA (c) methods; in all calculations the exocyclic hydroxymethylene group was assigned to the GT conformation. Contours are drawn at 1 kcal/mol intervals above the absolute minima obtained from each method, i.e. (a) Φ, Ψ = 60°, 0°; (b) Φ, Ψ = 180°, 0°; (c) Φ, Ψ = 60°, 0°.

contour map in Figure 1a is very akin to corresponding maps generated using MM2 [51]. The corresponding map calculated using the all atom CHARMm force field (cf. Figure 1b) displays similar features but suggests other minima (due to intramolecular hydrogen bonding). Contrastingly, contour maps calculated [52] using rigid-residue models like the one implemented in the HSEA approach usually show much smaller accessible (Φ, Ψ)-areas (see Figure 1c): when moving away from the



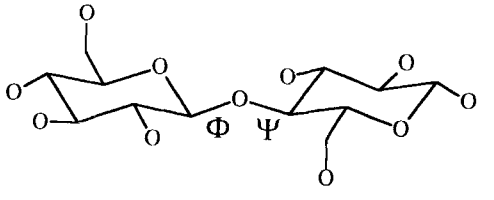


Figure 1 *continued*

local minimum energy locus, the energy of the system increases quickly to values of 50 kcal/mol and more [53]. The inappropriateness of the latter methods has been discussed extensively [54] and is evidenced by their starting model dependency and by their overestimation of energy barriers between (local) energy minima.

We have also carried out conformational analyses of the 11 possible dimannoses and two trimannoses [55]. The results from that study confirm the conclusions drawn above: conformational maps calculated using CHEAT are directly comparable with high-quality adiabatic MM2-type calculations that require extensive conformational searching at substantially higher computational expenses.

Molecular Dynamics of Glucose

Molecular dynamics calculations provide explicit information about the flexibility of molecules for a given potential function. Therefore, as a test of the suitability of the CHEAT force field for representing molecular motions and dynamical fluctuations, we carried out four 1000 ps molecular dynamics simulations using the α- and β-anomers of glucose in both the $^4C_1$ and $^1C_4$ conformations as starting

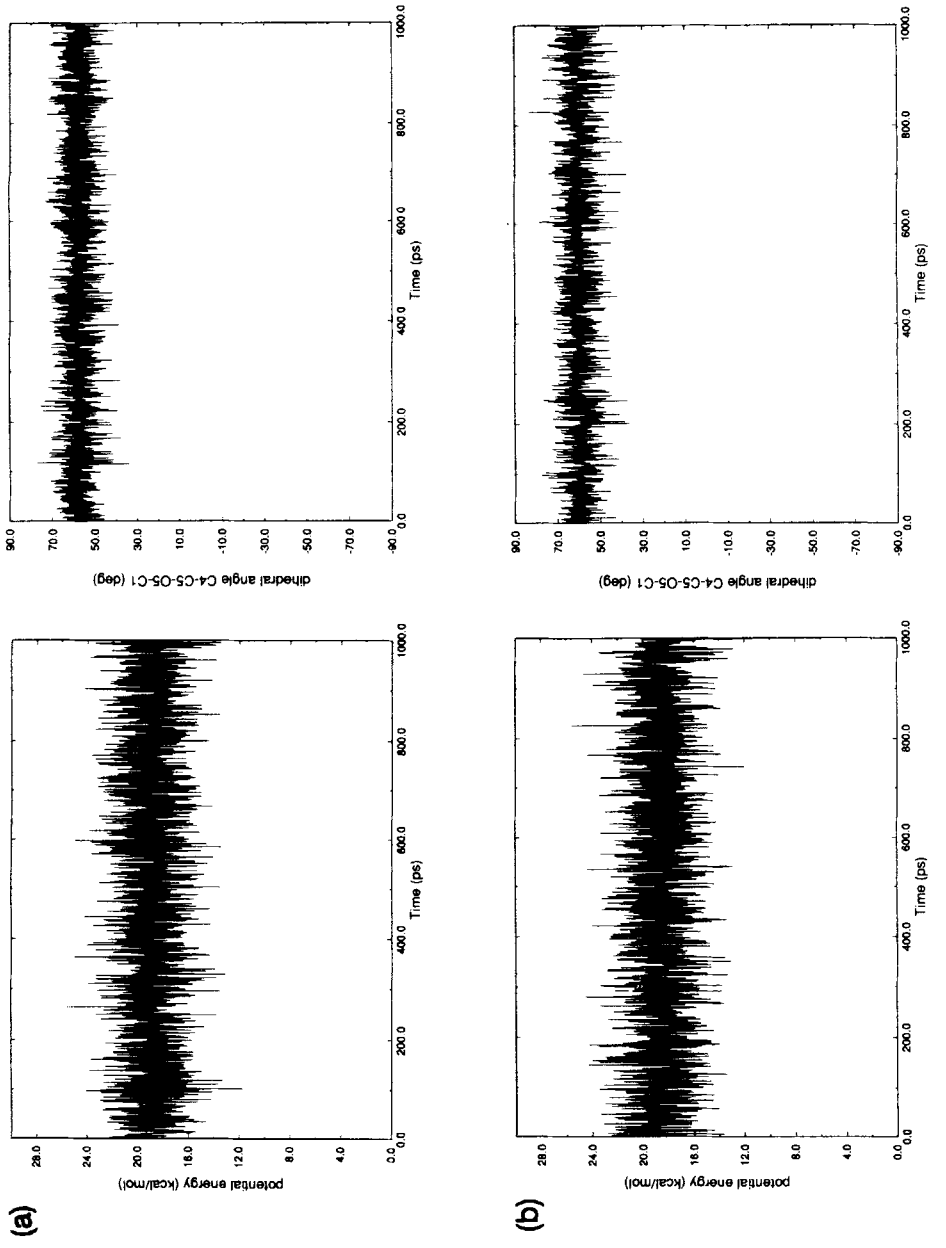


Figure 2 Time courses (1000 ps) for the molecular dynamics simulations of α- and β-glucose starting in the $^4C_1$ conformation (Fig. 2a and 2b, respectively) and the $^1C_4$ conformation (Fig. 2c and 2d, respectively): potential energy trajectories (lefthand side) and histories of the endocyclic C4-C5-O5-C1 torsion angle (righthand side).

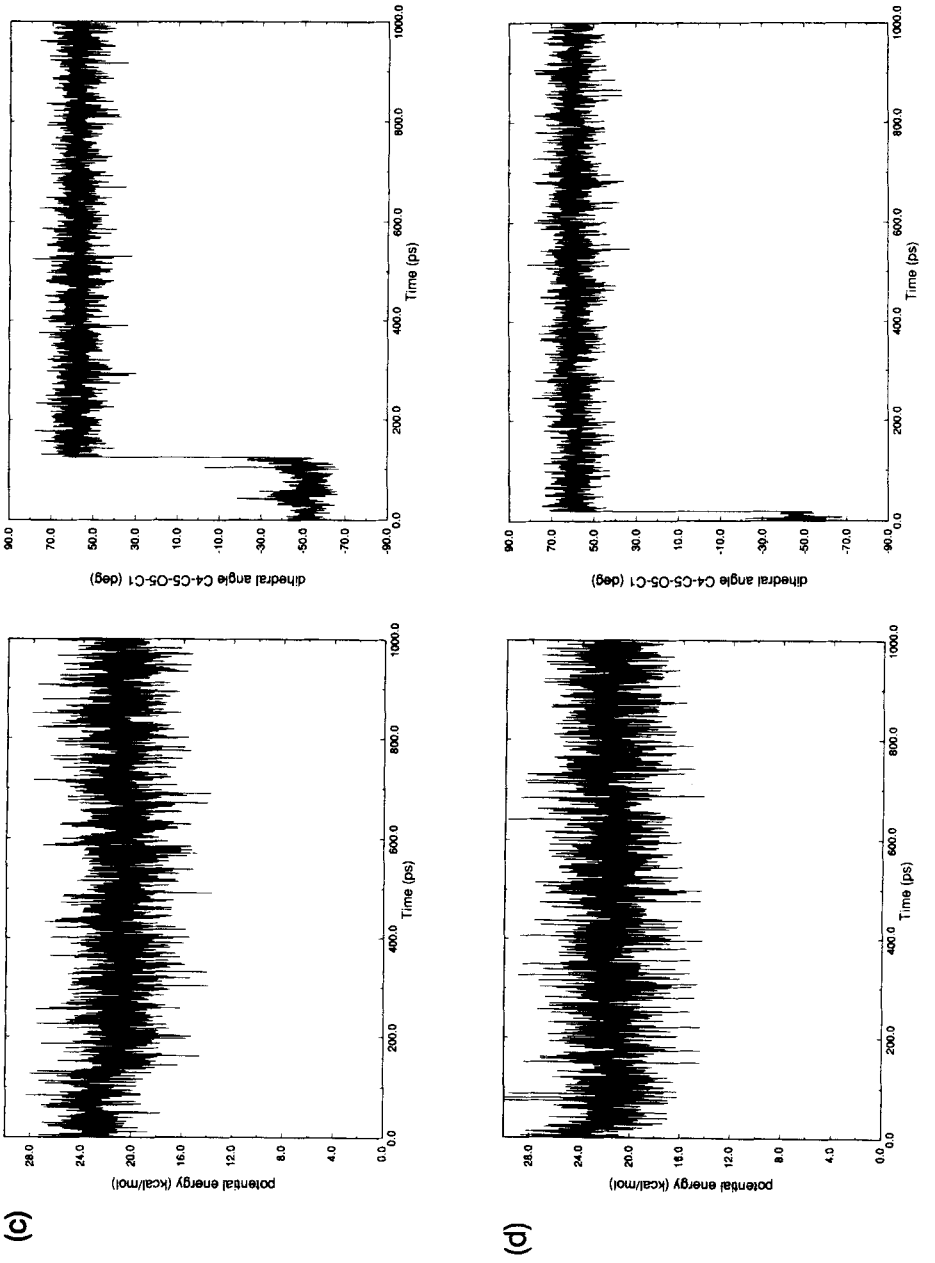


Figure 2 *continued*

structures. Molecular dynamics calculations for these four conformations using Rasmussen's PEF422 force field [8] have been reported previously by Brady [56,57]. Although the time span of the MD trajectories studied by Brady was considerably shorter than those presented here, a remarkable conformational flexibility was noted for these molecules. Moreover, conformational transitions between the two possible chair conformations were observed which appeared to take place *via* reasonably long-lived (viz., picosecond time scale) boat-intermediates.

The MD simulations carried out for the four ($^4C_1$ α-, $^4C_1$ β-, $^1C_4$ α- and $^1C_4$ β-) glucopyranose conformers using the CHEAT force field display conformational flexibilities similar to those reported by Brady. This is illustrated by Figure 2 which shows the history of the potential energies as well as the magnitude of the C4-C5-O5-C1 torsions in the calculated MD trajectories. In our calculations chair-to-chair conformational transitions only occurred when starting the simulations from the energetically unfavourable $^1C_4$ conformation. From Figure 2 it can be gleaned that in the latter simulations a transition to the favoured $^4C_1$ conformation is observed after approximately 125 ps and 20 ps for the α- and β-anomer, respectively. These conformational transitions take place in an abrupt manner: no intermediate conformations (skew-boat, half-chair, etc.) are observed for longer than a few fs.

Another interesting parameter to monitor is the relative population of the three staggered minima of the exocyclic hydroxymethyl group. It is known from NMR studies [58] that in aqueous solution the gauche-gauche (GG) rotamer is preponderant; the gauche-trans (GT) rotamer is considerably less populated and the trans-gauche (TG) rotamer is almost absent under these conditions. So far, this observation could not be reproduced by MD calculations; in fact, Brady found [56,57,59] that the TG rotamer is consistently the dominant conformer even when solvent molecules are explicitly taken into account [59]. Figure 3 displays the histories of the exocyclic C4-C5-C6-O6 torsion angle in the CHEAT molecular dynamics trajectories starting from $^4C_1$ α-, $^4C_1$ β-, $^1C_4$ α-, and $^1C_4$ β-glucopyranose, respectively. Scrutiny of the data in Figure 3 shows that the TG rotamer is sampled least frequently; the GT and GG rotamers are observed most (90%) of the time. When the data from all four molecular dynamics runs are combined the ratio TG:GT:GG amounts to 10:65:25. This ratio roughly corresponds to the relative potential energies for the three rotamers e.g. 6.29, 5.12 and 5.81 kcal/mol for β-glucose in the $^4C_1$ conformation. However, in contrast with experiment the GG rotamer is calculated to be less populated than the GT rotamer. It is suspected that this disagreement with experiment finds its origin in an inconsistency of the nonbonded interaction force field parameters. Further investigations towards this problem are in progress.

CONCLUSIONS

The CHEAT approach presented in this paper turns out to be a pragmatic, fairly general way to deal with the intra- vs. intermolecular hydrogen bonding problem that complicates carbohydrate modelling. From the above described results it can be concluded that our primary goal has been achieved, viz. the development of a carbohydrate force field that yields reliable, aqueous state compatible energetics in combination with reasonable geometries from *in vacuo* calculations. However, the

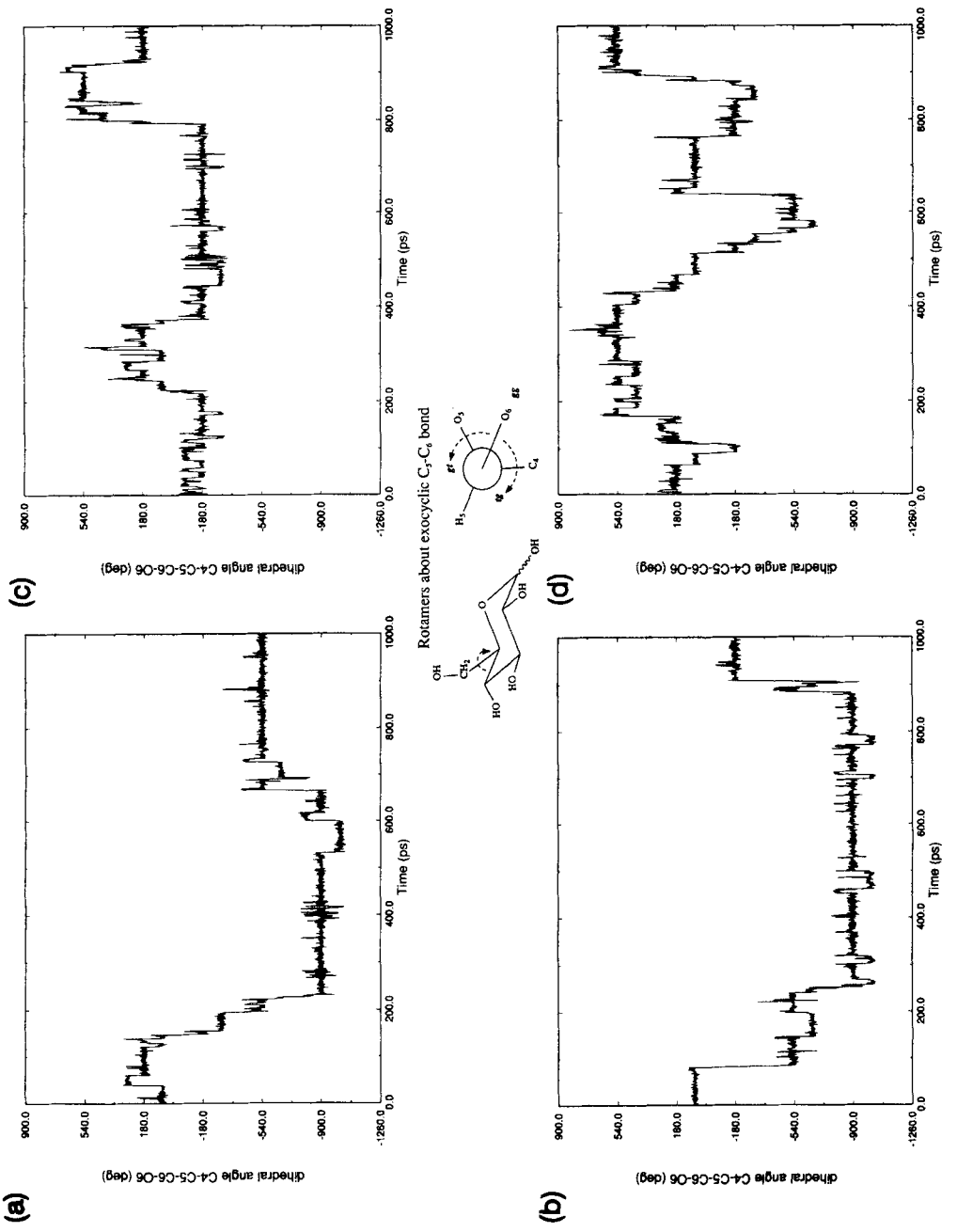


Figure 3 History (1000 ps) of the exocyclic C4-C5-C6-O6 dihedral angle for the molecular dynamics trajectories of α-glucose and β-glucose starting from the $^4C_1$ conformation (Fig. 3a and 3b, respectively) and the $^1C_4$ conformation (Fig. 3c and 3d, respectively).

second important objective, i.e. to leave the underlying CHARMm force field intact, could not be fully attained because of some fundamental problems with this force field. At this point it is good to note that preliminary attempts to apply the CHEAT approach to furanoses failed, presumably also because of the poor quality of the basic force field to describe small ring systems [22]. Notwithstandingly, the usefulness of the CHEAT approach is obvious from the conformational analyses and molecular dynamics examples presented for several saccharides. We expect that the CHEAT force field can also straightforwardly be implemented into other general force fields like GROMOS and AMBER. As a consequence, calculations using the CHEAT method are easy to set up and analyse because they can take advantage of the state-of-the-art graphics interfaces to e.g. CHARMm, GROMOS, etc., allowing the full functionality of those packages including molecular dynamics, docking, statistical analyses, and so forth.

The CHEAT approach has substantial advantages over established methods such as the HSEA or MM2-type approaches. The most important one is that CHEAT results in an improved energetical and conformational description of pyranoses as is evident from the cellobiose example. The inter-moiety hydrogen bonding that is unlikely to occur in water and in some cases leads to distorted conformational maps, is banned efficiently in CHEAT. Furthermore, the CHEAT approach is computationally very efficient, facilitating nanosecond range MD calculations of the dynamical conformational characteristics of "solute" carbohydrate molecules; the resulting ensembles are thus large enough to preclude "statistical accidents" [59].

Of course, there are also problems and limitations to the CHEAT approach. Because the carbohydrates modelled with CHEAT are not able to act as hydrogen bonding donors [60], CHEAT cannot be used for modelling protein-carbohydrate interactions. However, in the way we envisage such studies, CHEAT should be used to sample the conformational space of the uncomplexed carbohydrate. Subsequently, an all-atom force field may be applied for studying the interactions with the target protein. With respect to this last issue it is very important that the CHEAT parameter set is in balance with the basic force field parameters. As stated above, we did not quite succeed in leaving the underlying CHARMm force field completely intact. However, we hope that in future (improved) versions of this or another general valence force field, CHEAT can be implemented in such a way that it is a true superset of the basic force field.

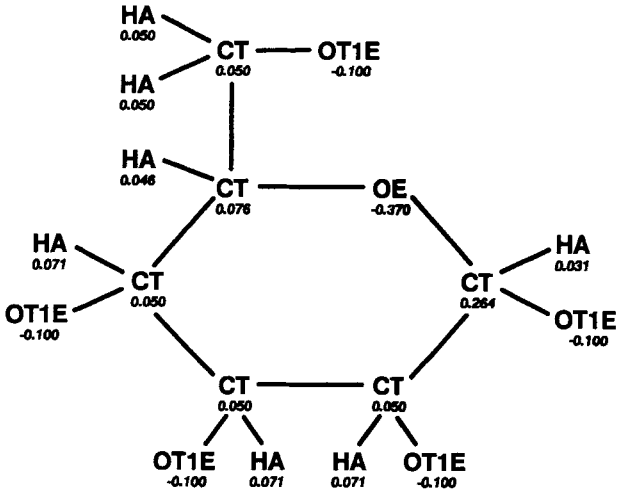
ACKNOWLEDGEMENT

The authors thank Mr. R. Schrijvers for his assistance in collecting the data needed for the evaluation of the pyranose ring conformations and Dr. L.M.J. Kroon-Batenburg for supplying the HSEA energy data of cellobiose.

APPENDIX

In this Appendix the parameters and data that together form the CHEAT force field are described. Since we used CHARMm version 21.2 for implementing the CHEAT approach, the potential energy function of this force field was applied. Further details on the minimisation protocols, etc. are found in the section Com-

putational Details. In Table 4 the bonded and nonbonded parameters from the CHEAT and CHARMm force fields needed for the simulation of pyranoses are collected. The atom types and electrostatic point charges for a CHEAT pyranose are summarised in scheme 1. Note that due to rounding errors the sum of the electrostatic point charges amounts to 0.06 instead of the theoretically required 0.0. Scheme 2 gives a representative example of a pyranose CHARMm residue topology file (RTF) to be used in CHEAT-type calculations.



Scheme 1

Table 4 Bonded and nonbonded parameters of the CHEAT force field needed to simulate pyranoses.

Bond lengths	k_r (kcal/mol.Å^2)	r_0(Å)	
CT-CT	268.0	1.515	[a]
CT-OE	380.0	1.40	[a]
CT-OT1E	320.0	1.43	
HA-CT	340.0	1.09	[a]
Bond angles	k_Θ (kcal/mol.rad^2)	Θ_0(°)	
CT-CT-CT	26.0	108.95	[a]
CT-CT-OE	80.0	109.47	[a]
CT-CT-OT1E	55.0	110.50	
CT-OE-CT	61.0	107.00	[a]
HA-CT-CT	37.5	109.47	[a]
HA-CT-HA	33.0	109.80	[a]
HA-CT-OE	45.5	109.47	[a]
HA-CT-OT1E	25.0	106.80	
OE-CT-OE	80.0	112.00	[a]
OE-CT-OT1E	80.0	109.47	

Table 4 *continued*

Dihedral angles	k_ϕ (kcal/mol)	periodicity n	δ(°)	
OT1E-CT-CT-OT1E	0.2	2	0.0	
OT1E-CT-CT-OT1E	0.2	3	0.0	
OE-CT-CT-OT1E	0.2	2	0.0	
OE-CT-CT-OT1E	0.2	3	0.0	
OE-CT-CT-OE	0.2	2	0.0	
OE-CT-CT-OE	0.2	3	0.0	
X-CT-CT-X	1.0	3	0.0	
OT1E-CT-OE-CT	0.5	2	0.0	
OT1E-CT-OE-CT	0.3	3	0.0	
OE-CT-OE-CT	0.5	2	0.0	
OE-CT-OE-CT	0.3	3	0.0	
X-CT-OE-X	1.3	3	0.0	[a]

VanderWaals parameters	Emin (kcal/mol)	R*(Å)	Emin14 (kcal/mol)	R*14(Å)	
HA	−0.045	1.468	[b]		[a]
CT	−0.0903	2.150	−0.0903	1.80	
OE	−0.1591	1.600	[b]		[a]
OT1E	−0.010	2.250	−0.1	1.50	

[a] Taken from the original CHARMM 21.2 force field.
[b] The 1,4-nonbonded interactions between atoms with no explicit parameters are scaled down by a factor 0.5.


```
* Topology File for CHARMm V 21.2 (based on POLYSACH30.RTF)
* Using extended hydroxyl atoms
*
   20    1
MASS     3 HA     1.00800 ! Aliphatic or aromatic hydrogen
MASS    10 CT    12.01100 ! Aliphatic carbon (tetrahedral)
MASS    50 OE    15.99940 ! Ether oxygen / Acetal oxygen
MASS   212 OT1E  17.007   ! Extended Oxygen atom with one hydrogen

AUTOGEN ANGLES
DEFAULT   FIRST NONE    LAST NONE

RESI GLUE  0.00  ! Beta D Glucose - extended atom
GROUP
ATOM C1   CT  0.264  !
ATOM H1   HA  0.031  !     H61
ATOM C2   CT  0.05   !       \
ATOM H2   HA  0.071  !     H62-C6-O6
ATOM C3   CT  0.05   !         |
ATOM H3   HA  0.071  !     H4  C5----O5  H1
ATOM C4   CT  0.05   !      \ / \      \ /
ATOM H4   HA  0.071  !    O4-C4  H5    C1-O1
ATOM C5   CT  0.076  !        \        /
ATOM H5   HA  0.046  !         C3----C2
ATOM C6   CT  0.05   !        / \    / \
ATOM H61  HA  0.050  !       O3 H3 O2 H2
ATOM H62  HA  0.05   !
ATOM O5   OE   -0.37
ATOM O2   OT1E -0.10
ATOM O3   OT1E -0.10
ATOM O4   OT1E -0.10
ATOM O6   OT1E -0.10
```

Scheme 2

```
ATOM O1   OT1E -0.10
BOND C1   C2    C2   C3    C3  C4    C4  C5   C5  O5    O5  C1
BOND C1   H1    C1   O1    C2  O2    C3  H3   C3  O3
BOND C6   H61   C6   H62   C6  O6
BOND C4   H4    C4   O4    C5  H5    C5  C6   C2  H2
DIHE O5 C1 C2 O2     O5 C1 C2 O2
DIHE O2 C2 C3 O3     O2 C2 C3 O3
DIHE O3 C3 C4 O4     O3 C3 C4 O4
DIHE O4 C4 C5 O5     O4 C4 C5 O5
DIHE O6 C6 C5 O5     O6 C6 C5 O5
DIHE C4 C5 O5 C1
DIHE C5 O5 C1 O1     C5 O5 C1 O1
```

Scheme 2 *continued*

REFERENCES AND NOTES

[1] (a) J.R. Rasmussen, "Effect of glycosylation on protein function", *Curr. Opin. Struct. Bio.* **2**, 682 (1992); (b) T. W. Rademacher, R. B. Parekh and R. A. Dwek, "Glycobiology", *Ann. Rev. Biochem.* **57**, 785 (1985); (c) C.A.A. van Boeckel, P.D.J. Grootenhuis and C.A.G. Haasnoot, "Specificity in the recognition process between charged carbohydrates and proteins", *Trends Pharm. Sci.* **12**, 241 (1991); (d) I. Tvaroška. "Theoretical aspects of structure and conformation of oligosaccharides", *Curr. Opin. Struct. Biol.* **2**, 661 (1991)

[2] *Computer Modeling of Carbohydrate Molecules*, A. D. French and J. W. Brady, eds, ACS Symposium Series, American Chemical Society, Washington, DC, 1990.

[3] B. Meyer, "Conformational aspects of oligosaccharides", *Top. Curr. Chem.* **154**, 141 (1990).

[4] J. Koehler, "Molecular dynamics simulations of carbohydrates", *Top. Mol. Struct. Biol.* **16**, 69 (1991).

[5] J. Tvaroška, "Computational methods for studying oligo- and polysaccharide conformations", *Pure & Appl. Chem.*, **61**, 1201 (1989).

[6] H. Thogersen, R.U. Lemieux, K. Bock, and B. Meyer, "Further justification for the exo-anomeric effect. Conformational analysis based on nuclear magnetic resonance, spectroscopy of oligosaccharides", *Can. J. Chem.*, **60**, 44 (1982).

[7] I. Tvaroška and S. Pérez, "Conformational-energy calculations for oligosaccharides: a comparison of methods and a strategy of calculation", *Carbohydr. Res.* **149**, 389 (1986).

[8] K. Rasmussen, "Conformation and anomer ratio of D-glucopyranose in different potential energy functions", *Acta Chem. Scand.*, **A36**, 323 (1982).

[9] S.N. Ha, A. Giammona, M. Field and J.W. Brady, "A revised potential-energy surface for molecular mechanics studies of carbohydrates", *Carbohydr. Res.* **180**, 207 (1988).

[10] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, "CHARMM: a program for macromolecular energy, minimization and dynamics calculations", *J. Comput. Chem.* **4**, 187 (1983).

[11] G.A. Jeffrey and R. Taylor, "The application of molecular mechanics to the structures of carbohydrates", *J. Comput. Chem.*, **1**, 99 (1980).

[12] N.L. Allinger, "Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms", *J. Am. Chem. Soc.* **99**, 8127 (1977).

[13] S. W. Homans, "A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Manαl-3Manβ1-4G1cNAc", *Biochemistry* **29**, 9110 (1990).

[14] J.N. Scarsdale, P. Ram, J.H. Prestegard and R.K. Yu, "A molecular mechanics-NMR pseudo-energy approach to the solution conformation of glycolipids", *J. Comput. Chem.*, **9**, 133 (1988).

[15] S.J. Weiner, P.A. Kollman, D.T. Nguyen and D.A. Case, "An all atom force field for simulations of proteins and nucleic acids", *J. Comput. Chem.*, **7**, 230 (1986).

[16] R. Stuike-Prill and B. Meyer, "A new force-field program for the calculation of glycopeptides and its application to a heptacosapeptide-decasaccharide of immunoglobulin G_1. Importance of 1-6 glycosidic linkages in carbohydrate-peptide interactions", *Eur. J. Biochem.* **194**, 903 (1990).

[17] M. K. Dowd, P. J. Reilly and A. D. French, "Conformational analysis of trehalose disaccharides and analogues using MM3", *J. Comput. Chem.* **13**, 102 (1992).
[18] B. R. Leeflang, J.F.G. Vliegenthart, L.M.J. Kroon-Batenburg, B.P. van Eijck and J. Kroon, "A proton NMR and molecular dynamics study of intramolecular hydrogen bonds in methyl β-cellobioside", *Carbohydr. Res.* **230**, 41 (1992).
[19] A. Imberty, K. D. Hardman, J. P. Carver and S. Pérez, "Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A", *Glycobiology* **1**, 631 (1991).
[20] M. Clark, R.D. Cramer III and N. van Opdenbosch, "Validation of the general purpose Tripos 5.2 force field", *J. Comput. Chem.* **10**, 982 (1989).
[21] P. Kollman, "Grand challenges in computational science: simulations on biological macromolecules", *Future Generation Computer Systems* **5**, 207 (1989).
[22] P.D.J. Grootenhuis, C.A.G. Haasnoot, unpublished results.
[23] CHARMm denotes a version of CHARMM (cf. Ref. 10) implemented by Molecular Simulations, Inc.
[24] Ref. 2, p. 162.
[25] When limiting oneself to three staggered minima for each pendant hydroxyl group, the number of conformations to evaluate for a tetrasaccharide amounts to 3^{20} ($\approx 3.48\ 10^9$). Even if one could generate, optimise and evaluate 100 conformations per cpu second, it would still take nearly 10.000 cpu hours to perform a systematic conformational analysis.
[26] E.L. Eliel, N.L. Allinger, S.J. Angyal and G.A. Morrison, *Conformational Analysis*; American Chemical Society, Washington, DC, 1981, Chpt. 6, pp 351–432.
[27] C.V. Goebel, W.L. Dimpfl and D.A. Brant, "The conformational energy of maltose and amylose", *Macromolecules* **3**, 644 (1970).
[28] C. Hervé du Penhoat, A. Imberty, N. Roques, V. Michon, J. Mentech, G. Descotes and S. Pérez, "Conformational behavior of sucrose and its deoxy analog in water as determined by NMR and molecular modelling", *J. Am. Chem. Soc.* **113**, 3720 (1991).
[29] L. Poppe and H. van Halbeek, "The rigidity of sucrose: just an illusion?", *J. Am. Chem. Soc.* **114**, 1092 (1992).
[30] B. Adams and L. Lerner, "Observation of hydroxyl protons of sucrose in aqueous solution: no evidence for persistent intramolecular hydrogen bonds", *J. Am. Chem. Soc.* **114**, 4827 (1992).
[31] Ref. 2, p.4.
[32] W. F. van Gunsteren and H. J. C. Berendsen, "Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry", *Angew. Chem. Int. Ed. Engl.*, **29**, 992 (1990).
[33] For an example of an extensive molecular dynamics study of α-D-glucose surrounded by 207 water molecules see Ref [59]. In the latter paper the exocyclic hydroxymethyl group was reported to adopt the experimentally not observed trans-gauche conformation for the largest part of the trajectory.
[34] Apart from the degrees of freedom allowed to molecules during relaxation, this also discerns our approach from the one implemented in HSEA-type methods: notwithstanding the fact that HSEA methods do not explicitly include hydroxyl protons, a careful parametrisation of the balance between intra- and intermolecular hydrogenbonds as adverted here has not been established.
[35] J. T. Edward, "Stability of glycosides to acid hydrolysis. A conformational analysis", *Chem. Ind. (London)* 1102 (1955).
[36] G. A. Jeffrey and J. H. Yates, "Application of ab initio molecular orbital calculations to the structural moieties of carbohydrates", *J. Am. Chem. Soc.* **101**, 820 (1979).
[37] S. Wolfe, A. Rauk, L. M. Tel and I.G. Csizmadia, "Theoretical study of the Edward-Lemieux effect (the anomeric effect). Stereochemical requirements of adjacent electron pairs and polar bonds", *J. Chem. Soc. B*, 136 (1971):
[38] S. Wolfe, "Gauche effect. Stereochemical consequences of adjacent electron pairs and polar bonds", *Acc. Chem. Res.* **5**, 102 (1972).
[39] Quanta/CHARMm version 3.0, released June 1990, Molecular Simulations Inc., 200 Fifth Ave, Waltham, MA.
[40] W. F. van Gunsteren and H. J. C. Berendsen, "Algorithms for macromoleculer dynamics and constraint dynamics", *Mol. Phys.*, **34**, 1311 (1977).
[41] S. J. Angyal, "Zusammensetzung und Konformation von Zuckern in Lösung", *Angew. Chem.*, **81**, 172 (1969).
[42] K. G. R. Pachler and P.C. Wessels, "Rotational isomerism. X. Nuclear magnetic resonance study of 2-fluoro-ethanol and ethylene glycol", *J. Mol. Struct.*, **6**, 471 (1970).

[43] S. Maleknia, B. R. Friedman, N. Abedi and M. Schwartz, "Conformational energies of 2-chloroethanol end ethylene glycol in hydrogen bonding solvents", *Spectrosc. Lett.*, **13**, 777 (1980).
[44] If present, the conformation of the exocyclic hydroxymethylene group at C5 was set to GG for the, $^4C_1$- and to GT for the $^1C_4$ conformations.
[45] Calculated using Angyal's scheme [41]; this value is in good correspondence with the experimental data given by Eliel et al. [Ref. 26, p. 44] stating a range of 1.5–2.1 kcal/mol with a recommended value of 1.7 kcal/mol.
[46] J.P. Lowe, "Barriers to internal rotation about single bonds", *Prog. Phys. Org. Chem.* **6**, 1 (1968).
[47] F.H. Allen, O. Kennard and R. Taylor, "Systematic analysis of structural data as a research technique in organic chemistry", *Acc. Chem. Res.*, **16**, 146 (1983).
[48] F.H. Allen, S. Bellard; M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B.G. Hummelink-Peters, O. Kennard, W.D.S. Motherwell, J.R. Rodgers, and D.G. Watson, "The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information", *Acta Cryst.*, **B35**, 233 (1979).
[49] D. Cremer and J.A. Pople, "A general definition of ring puckering coordinates", *J. Am. Chem. Soc.*, **97**, 1354 (1975).
[50] For a discussion of this problem and a proposal to its solution, see: (a) I. Tvaroška, T. Kozár and M. Hricovíni, "Oligosaccharides in solution. Conformational analysis by NMR spectroscopy and calculation", in Ref. 2, Chpt 10, pp. 162-176; (b) T. Kozár, F. Petrák, Z. Gálová and I. Tvaroška, "RAMM - a new procedure for theoretical conformational analysis of carbohydrates", *Carbohydr. Res.* **204**, 27 (1990).
[51] A.D. French, V.H. Tran and S. Pérez, "Conformational analysis of a disaccharide (cellobiose) with the molecular mechanics program (MM2)", in Ref. 2, Chpt. 12, pp. 191-212.
[52] L.M.J. Kroon-Batenburg, private communication.
[53] D.A. Brant and M.D. Christ, "Realistic conformational modeling of carbohydrates. Applications and limitations in the context of carbohydrate-high polymers", in Ref. 2, Chpt. 4, pp. 42-68.
[54] A.D. French and J. W. Brady, "Computer modeling of carbohydrates. An introduction", in Ref. 2, Chpt. 1, pp. 1-19.
[55] R. Schrijvers, P. D. J. Grootenhuis and C. A. G. Haasnoot, to be published.
[56] J. W. Brady, "Molecular dynamics simulations of α-D-glucose", *J. Am. Chem. Soc.*, **108**, 8153 (1986).
[57] J. W. Brady, "Molecular dynamics simulations of β-D-glucopyranose", *Carbohydr. Res.*, **165**, 306 (1987).
[58] C. Altona and C.A.G. Haasnoot, "Prediction of anti and gauche vicinal proton-proton coupling constants for hexapyranose rings using a generalized Karplus equation", *Org. Magn. Reson.* **13**, 417 (1980).
[59] J.W. Brady, "Molecular dynamics simulations of α-D-glucose in aqueous solution", *J. Am. Chem. Soc.* **111**, 5155 (1989).
[60] Note that the OT1E atoms can still accept hydrogen bonds due to their (small) negative charge of −0.1.